Exhibit A

Reference Manual on Scientific Evidence

Third Edition

Committee on the Development of the Third Edition of the
Reference Manual on Scientific Evidence

Committee on Science, Technology, and Law
Policy and Global Affairs

FEDERAL JUDICIAL CENTER

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

C. Does the Sample Approximate the Relevant Characteristics of the Population?

Identification of a survey population must be followed by selection of a sample that accurately represents that population.[91] The use of probability sampling techniques maximizes both the representativeness of the survey results and the ability to assess the accuracy of estimates obtained from the survey.

Probability samples range from simple random samples to complex multistage sampling designs that use stratification, clustering of population elements into various groupings, or both. In all forms of probability sampling, each element in the relevant population has a known, nonzero probability of being included in the sample.[92] In simple random sampling, the most basic type of probability sampling, every element in the population has a known, equal probability of being included in the sample, and all possible samples of a given size are equally likely to be selected.[93] Other probability sampling techniques include (1) stratified random sampling, in which the researcher subdivides the population into mutually exclusive and exhaustive subpopulations, or strata, and then randomly selects samples from within these strata; and (2) cluster sampling, in which elements are sampled in groups or clusters, rather than on an individual basis.[94] Note that selection probabilities do not need to be the same for all population elements; however, if the probabilities are unequal, compensatory adjustments should be made in the analysis.

Probability sampling offers two important advantages over other types of sampling. First, the sample can provide an unbiased estimate that summarizes the responses of all persons in the population from which the sample was drawn; that is, the expected value of the sample estimate is the population value being estimated. Second, the researcher can calculate a confidence interval that describes explicitly how reliable the sample estimate of the population is likely to be. If the sample is unbiased, the difference between the estimate and the exact value is called the sampling error.[95] Thus, suppose a survey collected responses from a simple random sample of 400 dentists selected from the population of all dentists

91. MCL 4th, *supra* note 16, § 11.493. *See also* David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section II.B, in this manual.

92. The exception is that population elements omitted from the sampling frame have a zero probability of being sampled.

93. Systematic sampling, in which every *n*th unit in the population is sampled and the starting point is selected randomly, fulfills the first of these conditions. It does not fulfill the second, because no systematic sample can include elements adjacent to one another on the list of population members from which the sample is drawn. Except in unusual situations when periodicities occur, systematic samples and simple random samples generally produce the same results. Thomas Plazza, *Fundamentals of Applied Sampling*, *in* Handbook of Survey Research, *supra* note 1, at 139, 145.

94. *Id*. at 139, 150–63.

95. *See* David H. Kaye & David A. Freedman, *supra* note 91, Glossary, for a definition of sampling error.

licensed to practice in the United States and found that 80, or 20%, of them mistakenly believed that a new toothpaste, Goldgate, was manufactured by the makers of Colgate. A survey expert could properly compute a confidence interval around the 20% estimate obtained from this sample. If the survey were repeated a large number of times, and a 95% confidence interval was computed each time, 95% of the confidence intervals would include the actual percentage of dentists in the entire population who would believe that Goldgate was manufactured by the makers of Colgate.[96] In this example, the margin of error is ±4%, and so the confidence interval is the range between 16% and 24%, that is, the estimate (20%) plus or minus 4%.

All sample surveys produce estimates of population values, not exact measures of those values. Strictly speaking, the margin of error associated with the sample estimate assumes probability sampling. Assuming a probability sample, a confidence interval describes how stable the mean response in the sample is likely to be. The width of the confidence interval depends on three primary characteristics:

1. Size of the sample (the larger the sample, the narrower the interval);
2. Variability of the response being measured; and
3. Confidence level the researcher wants to have.[97]

Traditionally, scientists adopt the 95% level of confidence, which means that if 100 samples of the same size were drawn, the confidence interval expected for at least 95 of the samples would be expected to include the true population value.[98]

Stratified probability sampling can be used to obtain more precise response estimates by using what is known about characteristics of the population that are likely to be associated with the response being measured. Suppose, for example, we anticipated that more-experienced and less-experienced dentists might respond differently to Goldgate toothpaste, and we had information on the year in which each dentist in the population began practicing. By dividing the population of dentists into more- and less-experienced strata (e.g., in practice 15 years or more versus in practice less than 15 years) and then randomly sampling within experience stratum, we would be able to ensure that the sample contained precisely

96. Actually, because survey interviewers would be unable to locate some dentists and some dentists would be unwilling to participate in the survey, technically the population to which this sample would be projectable would be all dentists with current addresses who would be willing to participate in the survey if they were asked. The expert should be prepared to discuss possible sources of bias due to, for example, an address list that is not current.

97. When the sample design does not use a simple random sample, the confidence interval will be affected.

98. To increase the likelihood that the confidence interval contains the actual population value (e.g., from 95% to 99%) without increasing the sample size, the width of the confidence interval can be expanded. An increase in the confidence interval brings an increase in the confidence level. For further discussion of confidence intervals, see David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section IV.A, in this manual.

proportionate representation from each stratum, in this case, more- and less-experienced dentists. That is, if 60% of dentists were in practice 15 years or more, we could select 60% of the sample from the more-experienced stratum and 40% from the less-experienced stratum and be sure that the sample would have proportionate representation from each stratum, reducing the likely sampling error.[99]

In proportionate stratified probability sampling, as in simple random sampling, each individual member of the population has an equal chance of being selected. Stratified probability sampling can also disproportionately sample from different strata, a procedure that will produce more precise estimates if some strata are more heterogeneous than others on the measure of interest.[100] Disproportionate sampling may also used to enable the survey to provide separate estimates for particular subgroups. With disproportionate sampling, sampling weights must be used in the analysis to accurately describe the characteristics of the population as a whole.

Although probability sample surveys often are conducted in organizational settings and are the recommended sampling approach in academic and government publications on surveys, probability sample surveys can be expensive when in-person interviews are required, the target population is dispersed widely, or members of the target population are rare. A majority of the consumer surveys conducted for Lanham Act litigation present results from nonprobability convenience samples.[101] They are admitted into evidence based on the argument that nonprobability sampling is used widely in marketing research and that "results of these studies are used by major American companies in making decisions of considerable consequence."[102] Nonetheless, when respondents are not selected randomly from the relevant population, the expert should be prepared to justify the method used to select respondents. Special precautions are required to reduce the likelihood of biased samples.[103] In addition, quantitative values computed from such samples (e.g., percentage of respondents indicating confusion) should be viewed as rough

99. *See* Pharmacia Corp. v. Alcon Lab., 201 F. Supp. 2d 335, 365 (D.N.J. 2002).

100. Robert M. Groves et al., Survey Methodology, Stratification and Stratified Sampling, 106–18 (2004).

101. Jacob Jacoby & Amy H. Handlin, *Non-Probability Sampling Designs for Litigation Surveys*, 81 Trademark Rep. 169, 173 (1991). For probability surveys conducted in trademark cases, see James Burrough, Ltd. v. Sign of Beefeater, Inc., 540 F.2d 266 (7th Cir. 1976); Nightlight Systems, Inc., v. Nite Lights Franchise Sys., 2007 U.S. Dist. LEXIS 95565 (N.C. Ga. July 17, 2007); National Football League Props., Inc. v. Wichita Falls Sportswear, Inc., 532 F. Supp. 651 (W.D. Wash. 1982).

102. National Football League Props., Inc. v. New Jersey Giants, Inc., 637 F. Supp. 507, 515 (D.N.J. 1986). A survey of members of the Council of American Survey Research Organizations, the national trade association for commercial survey research firms in the United States, revealed that 95% of the in-person independent contacts in studies done in 1985 took place in malls or shopping centers. Jacoby & Handlin, *supra* note 101, at 172–73, 176. More recently, surveys conducted over the Internet have been administered to samples of respondents drawn from panels of volunteers; *see infra* Section IV.G.4 for a discussion of online surveys. Although panel members may be randomly selected from the panel population to complete the survey, the panel population itself is not usually the product of a random selection process.

103. *See infra* Sections III.D–E.

indicators rather than as precise quantitative estimates.[104] Confidence intervals technically should not be computed, although if the calculation shows a wide interval, that may be a useful indication of the limited value of the estimate.

D. What Is the Evidence That Nonresponse Did Not Bias the Results of the Survey?

Even when a sample is drawn randomly from a complete list of elements in the target population, responses or measures may be obtained on only part of the selected sample. If this lack of response is distributed randomly, valid inferences about the population can be drawn with assurance using the measures obtained from the available elements in the sample. The difficulty is that nonresponse often is not random, so that, for example, persons who are single typically have three times the "not at home" rate in U.S. Census Bureau surveys as do family members.[105] Efforts to increase response rates include making several attempts to contact potential respondents, sending advance letters,[106] and providing financial or nonmonetary incentives for participating in the survey.[107]

The key to evaluating the effect of nonresponse in a survey is to determine as much as possible the extent to which nonrespondents differ from the respondents in the nature of the responses they would provide if they were present in the sample. That is, the difficult question to address is the extent to which nonresponse has biased the pattern of responses by undermining the representativeness of the sample and, if it has, the direction of that bias. It is incumbent on the expert presenting the survey results to analyze the level and sources of nonresponse, and to assess how that nonresponse is likely to have affected the results. On some occasions, it may be possible to anticipate systematic patterns of nonresponse. For example, a survey that targets a population of professionals may encounter difficulty in obtaining the same level of participation from individuals with high-volume practices that can be obtained from those with lower-volume practices. To enable the researcher to assess whether response rate varies with the volume of practice, it may be possible to identify in advance potential respondents

104. The court in Kinetic Concept, Inc. v. Bluesky Medical Corp., 2006 U.S. Dist. LEXIS 60187, *14 (W.D. Tex. Aug. 11, 2006), found the plaintiff's survey using a nonprobability sample to be admissible and permitted the plaintiff's expert to present results from a survey using a convenience sample. The court then assisted the jury by providing an instruction on the differences between probability and convenience samples and the estimates obtained from each.

105. 2 Gastwirth, *supra* note 64, at 501. This volume contains a useful discussion of sampling, along with a set of examples. *Id.* at 467.

106. Edith De Leeuw et al., *The Influence of Advance Letters on Response in Telephone Surveys: A Meta-analysis*, 71 Pub. Op. Q. 413 (2007) (advance letters effective in increasing response rates in telephone as well as mail and face-to-face surveys).

107. Erica Ryu et al., *Survey Incentives: Cash vs. In-kind; Face-to-Face vs. Mail; Response Rate vs. Nonresponse Error*, 18 Int'l J. Pub. Op. Res. 89 (2005).